Visual Informatics 3 (2019) 38–47

Contents lists available at ScienceDirect

Visual Informatics

journal homepage: www.elsevier.com/locate/visinf

Exploration behavior of group-in-a-box layouts

Yuki Ueno [a], Hiroaki Natsukawa [b,*], Nozomi Aoyama [a], Koji Koyamada [b]

[a] *Department of Electrical Engineering, Kyoto University, Kyoto, Japan*
[b] *Academic Center for Computing and Media Studies, Kyoto University, Kyoto, Japan*

ARTICLE INFO

Article history:
Available online 27 March 2019

Keywords:
Visual search
Eye-tracking
Group-in-a-box layout

ABSTRACT

To improve visualization, it is necessary to optimize the design by analyzing the behavior of users as well as improving the evaluation index of the computational experiment and the task performance (e.g., the correct answer rate and completion time) in the user experiment. Although various studies have investigated the influence of user behavior on the evaluation of visualization, majority of these studies focused on simple visualization tasks. A simple task does not indicate a simple visualization comprising a few visualization elements but a task in which the information obtained from visualization is the only clue for completing the task. However, a few studies have targeted complicated tasks in which multiple information obtained from visualization is considered to be a clue for completing the task regardless of the number of elements that are contained in the visualization. Therefore, in this study, we investigated the behavior of the participants who have performed complicated tasks. We selected two types of group-in-a-box (GIB) layouts, which can be considered to be a complicated visualization method, as the target of the user experiment. In the user experiment, participants were asked to perform an exploration task specific to GIB layouts; which group has the maximum number of intra-edges? We also collected the eye-tracking data in addition to task performance. The results showed that the correct answer rate is considerably affected by the visualization factor; whether the correct answer, the box with maximum number of intra-edges, is the box with the largest area. Furthermore, an analysis of the collected eye-tracking data revealed that this visualization factor affected the exploration behavior of the participants; however, it did not affect the location at which the participants were focused on. The obtained results indicated that the visualization elements that were not considered by the visualization designer can influence the task of extracting information from the data. Therefore, designers have to configure the visualization by considering the visual cognitive behavior of the users.

1. Introduction

The visualization techniques make it possible to recognize data, understand data, and generate new insights. Therefore, the interactions between humans and visualization techniques are considered to be important, and it is necessary to incorporate humans as a part of the visualization system during the design and evaluation of visualization techniques. Further, to effectively utilize visualization, it is not only necessary to improve the evaluation index of the computational experiment and the task performance of the user experiment (e.g., the correct answer rate and completion time) but also to analyze the human behavior in the user experiment and to optimize the design while considering the experimental results. This promotes the better usage of visualization.

There are several conventional methods, such as thinking aloud protocols and interaction logs, for recording the human behavior. In recent years, various gaze measurement techniques have been developed, which makes it possible to quantitatively easily record the user behavior without the requirement of any specialized knowledge. Thus, in the field of visualization, these methods have been adopted for evaluation of visualization. For example, it has been used to evaluate the visualization methods supporting the multi-attribute decision making (Kim et al., 2012) and node-link diagrams (Burch et al., 2011; Netzel et al., 2014). In addition, new evaluation methods have been proposed by combining the thinking aloud protocols, interaction logs, and eye-tracking (Blascheck et al., 2016).

Several previous studies have focused on analyzing the user behavior as a part of visualization evaluation even though several evaluation tasks were considered to be simple. A simple task does not mean a task of simple visualization comprising a few visualization elements but a task in which the information obtained from visualization is the only clue to complete the task. However,

* Corresponding author.
E-mail address: natsukawa.hiroaki.3u@kyoto-u.ac.jp (H. Natsukawa).

Peer review under responsibility of Zhejiang University and Zhejiang University Press.

https://doi.org/10.1016/j.visinf.2019.03.005
2468-502X/

a complicated task means a task in which multiple information obtained from the visualization is a clue to complete the task regardless of the number of elements comprised in the visualization. To the best of our knowledge, several previous studies have focused on analyzing the exploration behavior of the participants while performing simple tasks (Burch et al., 2011; Netzel et al., 2014; Huang et al., 2008); however, only a few studies have focused on analyzing the user behavior while performing a complicated task. Further, the design guidelines can be improved by adding human cognitive processes and behavioral patterns. Generally, it is necessary to conduct a behavior analysis related to the complicated visualization evaluation task to improve the design of the visualization method.

Based on the aforementioned background, this study focused on the analysis of users' behavior while extracting the information from a complicated task. The group-in-a-box (GIB) layout was selected as the target of this experiment. The GIB layout is an efficient graph-drawing method that is designed to visualize the group structure of graphs (Rodrigues et al., 2011; Chaturvedi et al., 2014; Onoue and Koyamada, 2017). Various types of GIB layouts have been proposed, and computational experiments have been conducted to compare these layouts (Chaturvedi et al., 2014; Onoue and Koyamada, 2017). The layouts basically comprise nodes, edges, and boxes, which surround the nodes belonging to the same group. In addition to being composed of multiple elements, the size of the box changes in proportion to the number of nodes in the group, resulting in a change in the appearance of the edge-connecting nodes belonging to the same group. Further, we can obtain plenty of information from the GIB layout; therefore, it is possible that we can observe multiple exploration behaviors depending on the characteristics of data.

In the user experiment, we obtained the eye-tracking data from six participants who were asked to perform an exploration task using GIB layouts. We selected the force-directed GIB (FD-GIB) and tree-reordered GIB (TR-GIB) as the user experiment targets considering the results of the experiment 1. In addition, we also selected an evaluation task (i.e., which group has the maximum number of intra-edges?) for these GIB layouts based on the results of the experiment 1. While the participants were performing the task, we recorded their performance, including their completion time and the correct answer rate. Additionally, their eye movements were also recorded.

The analysis of task performance revealed that the correct answer rate is considerably affected by whether the correct answer is the box with the largest area. By analyzing the eye-tracking data, we confirmed that the exploration behavior in both FD-GIB and TR-GIB changed depending on the visualization factor affecting the correct answer rate. Further, the fixation point of the boxes did not change because of the visualization factor.

The primary contributions of this study can be summarized as follows.

- We discovered that the visualization factor affected the task performance.
- We also found that the exploration behavior can be changed by the visualization factor, which affected the task performance.

2. Related work

The analysis of the user behavior while using visualization systems has been investigated in several studies (Pohl et al., 2009; Burch et al., 2011; Netzel et al., 2017, 2014; Kim et al., 2012). In these studies, the eye-tracking system was used to record the users' eye movement. The eye-tracking data made it possible to understand the manner in which the users used the developed visualization system and provided insights into the users' reasoning methods and problem-solving strategies (Andrienko et al., 2012). Therefore, eye-tracking helps improve the visualization system by evaluating the usefulness and readability of the visualization technology with respect to visual cognition.

For instance, Netzel et al. (2017) evaluated four variants of geographic map annotation (the within-image annotation, grid reference annotation, directional annotation, and miniature annotation). The participants were asked to find the specified label within the map as fast and as accurately as possible. While they were performing the task, the eye-tracking data and completion time of the participants were recorded. The obtained result denotes that the within-image annotation was outperformed by all the remaining annotation methods. Miniature annotation resulted in optimal completion time. In addition, the eye-tracking data revealed that the participants used different task strategies for different geographic map annotations. Burch et al. (2011) explored three types of tree diagrams: a traditional tree layout, an orthogonal tree layout, and a radial tree layout. Participants were asked to search for the least common ancestor for a given set of marked leaf nodes, which can be considered to be a typical hierarchical exploration task. At that time, the eye-tracking data was recorded using the eye-tracker in addition to the correct answer rate and the completion time of the task. It is clear from the eye-tracking data that the exploration strategies are different for each method. The participants frequently cross-checked their task solutions and required more time to complete the task while using the radial layout than while using the remaining layouts.

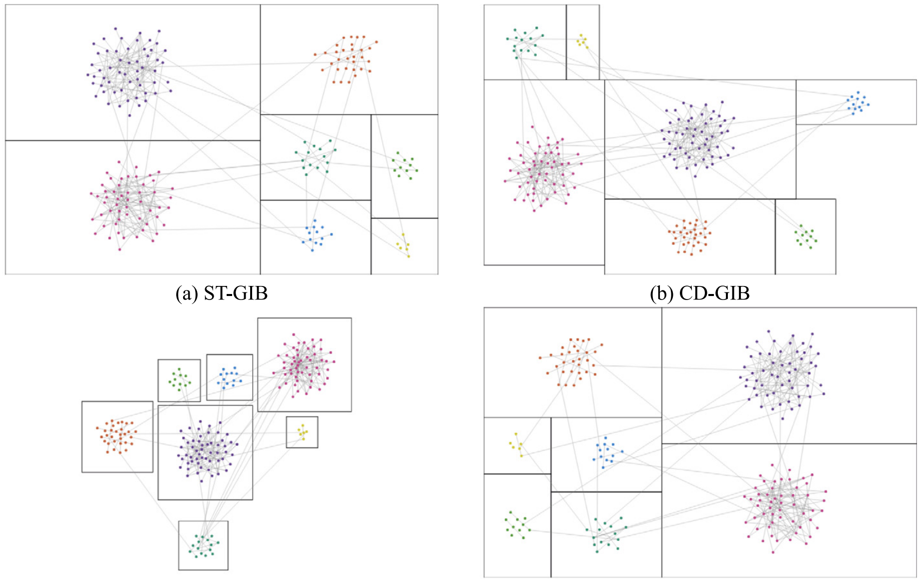
Although several studies related to the analysis of the task solving behavior have been conducted in the past, the majority of those studies targeted simple evaluation tasks. Hence, in this study, we selected a more complicated task as the target of the user experiment and investigated the task solving behavior based on visual cognition.

3. Experiment 1

We conducted a first experiment with 20 participants to determine the GIB layout and the evaluation task that are appropriate for the experiment 2. In the experiment 2, we intend to design a complicated task in which multiple pieces of information obtained from GIB layout can be used as a clue for completing the task.

3.1. GIB Layout

The GIB layout is a graph-drawing method designed to visualize the group structure of graphs (Rodrigues et al., 2011; Chaturvedi et al., 2014; Onoue and Koyamada, 2017). In GIB, all the nodes in the group are placed within a box whose size is proportional to the number of nodes. Therefore, while using GIB, it is possible to simultaneously visualize the group structure, the relation between various groups, and the size of the groups in the graph. In this study, we selected the GIB layout as the evaluation target because of two reasons. First, GIB layouts comprise various kinds of visualization elements, and we can obtain multiple pieces of information from the layout. Hence, using GIB layouts, multiple pieces of information can be obtained and used as a clue to achieve the task, and multiple exploration behaviors are observed to exist. Second, GIB layouts are suitable for performing eye-tracking analyses based on the area of interest (AOI). To be specific, in GIB layouts, the screen is divided into boxes, and the boxes can be regarded as the AOIs. There are several forms of GIB layouts; subsequently, we will present the four evaluated GIB layouts, which are ST-GIB, CD-GIB, FD-GIB, and TR-GIB. Examples of these layouts are presented in Fig. 1.



(a) ST-GIB
(b) CD-GIB
(c) FD-GIB
(d) TR-GIB

Fig. 1. Examples of group-in-a-box (GIB) layouts.

ST-GIB Squarified treemap GIB (ST-GIB), which is based on squarified treemaps proposed by Bruls et al. (2000), were developed by Rodrigues et al. (2011). In ST-GIB, groups are denoted in the shape of tiles that contain nodes belonging to the groups. This method can be used to easily fill the space using low aspect-ratio boxes, which is an important factor while analyzing the rectangular content (Bruls et al., 2000). However, this method does not consider the relation among the nodes when the boxes are arranged; therefore, it includes edge crossing, which tends to hamper the user's understanding of the depicted networks (Becker et al., 1995; Purchase, 1997, 1998; Purchase et al., 2002).

CD-GIB Chaturvedi et al. (2014) developed croissant-and-doughnut GIB (CD-GIB). They improved ST-GIB with respect to the number of edges between the groups, allowing the network information to be considered. Further, the boxes are placed according to their G-degree and G-skewness. A group's G-degree can be defined as the number of other groups to which the group is connected, and the G-skewness refers to the fraction of nodes present in the two most-connected groups (the two groups that have the highest G-degree). Based on the G-degree and G-skewness, a layout is chosen from among the croissant-GIB, doughnut-GIB, and ST-GIB layouts. We used the criteria defined by Chaturvedi et al. to select a layout from among the three layouts. The doughnut-GIB places the most connected group in the center of the screen. The other groups are arranged around them, so this layout looks doughnut shape. On the other hand, the croissant-GIB places the most connected group in the center top of the screen. The other groups are arranged around them, so this layout looks croissant shape. Using these layouts, the most connected box with the highest G-degree is placed close to the center. Therefore, the number of edge crossings is small, and the readability is expected to be better than that of ST-GIB. Regardless, the aspect ratio tends to become worse, and there is a possibility of the readability being affected (Bruls et al., 2000).

FD-GIB Force-directed GIB (FD-GIB) was also developed by Chaturvedi et al. (2014). This layout uses a force-directed layout to arrange each box according to their attraction to the center and the repulsion between boxes. Because this layout can create overlaps, we removed such overlaps using the PRISM method (Gansner and Hu, 2008). Although this layout is suitable for depicting the topology of an entire network, it may be difficult to understand the relations that exist in a single group because each box can only occupy a small area. However, in this layout, the aspect ratio of each box can be made constant; therefore, we expect the users to be easily able to compare the box sizes.

TR-GIB Onoue and Koyamada (2017) proposed tree-reordered GIB (TR-GIB). This layout is based on ST-GIB and is optimized so that the lengths of all edges of the ST-GIB are minimized. More specifically, this layout minimizes the weighted sum of the distances between groups by reordering the sibling nodes in the ST-GIB layout. Because this layout is optimized to minimize the distance between groups, it has fewer edge crossings than in the case of ST-GIB. Hence, this layout is expected to exhibit the advantages of ST-GIB's good aspect ratio and effective use of the screen as well as the advantage of having less edge crossings.

3.2. Tasks and stimuli

We designed the evaluation tasks according to the method proposed by Vehlow et al. (2017) and Saket et al. (2014). Vehlow et al. provided four task taxonomies to evaluate the clustered graph visualizations: group-only tasks (GOT), group vertex tasks (GVT), group edge tasks (GET), and group network tasks (GNT). Further, GOT, GVT, and GNT can be used for performing the GIB evaluations; however, GET, which is used for the networks whose edges are grouped, cannot be used for performing the GIB evaluations. Saket et al. also provided several types of tasks. Both these research groups showed several examples of each task type. We selected the four tasks discussed below from those examples.

Task 1(GOT): How many groups are present in this graph?

Task 2(GVT): Which group has the maximum (or minimum) number of nodes?

Task 3(Intra-GNT): Which group has the maximum (or minimum) number of intra-edges?

Task 4(Inter-GNT): Which group has the maximum number of inter-edges?

The GIB layout is a visualization method used to visualize the relations between groups and within a group. Therefore, GIB can be used to denote both group intra-edges and group inter-edges. Based on this feature, the following two types of GNT tasks have been selected: intra-GNT and inter-GNT. The group intra-edges connect a node in one group to another node in the same group. Conversely, the group inter-edges connect one node in a group to a node in another group. In addition, GIB should exhibit good performance while showing the structure of both small and large boxes. Therefore, in tasks 2 and 3, we designed two tasks for each situation and changed the question (i.e., maximum or minimum) for half of the tasks.

3.3. Data and layout generation

In this experiment, random data were generated using the same method as the method used by Onoue and Koyamada (2017), and all the parameters related to the data generation were defined in a similar manner. However, there are two differences between our method and the method proposed by Onoue et al. The first difference is in the procedure that is used to randomly set the number of groups based on a normal distribution with m_{mean} and m_{stdev} ranging from m_{min} to m_{max}. The second difference is in the procedure that is used for setting the number of vertices in a vertex set V_i, which can be determined based on the random numbers that follow a normal distribution with v_{mean}, v_{stdev}, and v_{min}. Further, we calibrated each parameter to ensure that our data will closely resemble the Twitter data used by Chaturvedi et al. (2014). First, we calibrated the parameters to ensure that the number of groups, nodes, and edges correspond to the number of those used by Chaturvedi et al. Second, we reduced the number of nodes and edges by multiplying v_{mean} and v_{stdev} with 0.4 and by multiplying p_{in}, p_{bridge} and p_{out} with 0.3 because the number of nodes and edges were considerably large to understand. The parameters used are presented in Table 1.

The generated data were further visualized by applying each of the GIB layouts. Further, the GIB layouts can only arrange the boxes; therefore, it is necessary to determine the coordinates of nodes in each box. We selected a force layout from among the layouts available for arranging the nodes in a box. In this method, the nodes are arranged according to the repulsion between the nodes, the attraction between adjacent nodes, and the gravity level from the center of the group tile to which they belong. Although the layout method in the box affects the results of the task, the force layout is adopted in this study because it is known to reduce edge crossing and increase readability (Kobourov, 2004).

3.4. Participants

We used a within-subject study design with 20 participants, including 12 males and 8 females (age range: 18–24 years and mean: 20.8 years). Participants were not engaged in the visualization study but had adequate literacy for extracting information from diagrams and tables. All the participants were not familiar with the GIB layouts, and all of them either had normal or corrected-to-normal color vision. At the beginning of the experiment, all the participants signed an informed consent form. Further, each participant was compensated with 3000 yen.

3.5. Study procedure

The experiments required 1.5–2 h, including the preparation, explanation and breaks. In the experiment, participants had to perform four types of tasks for four different GIB layouts. At the beginning of the experiment, participants filled the informed consent form and the basic information questionnaire (name, age, and gender). Subsequently, we explained each GIB layout to the participants who were sitting at a position that was 65 cm from the screen. Finally, we explained the task in detail, which was followed by a tutorial, to ensure that the participants understood the network and to ensure that they are able to complete the tasks. Although the trial data provided in the tutorial was different from those provided in the actual experiments, the tutorial provided a sufficient practical guide to the experimental procedures to the participants. After completing this step, the participants began the actual experiment. In the actual experiment, four types of tasks were set for the four different GIB layouts, and 30 trials were prepared for each type. Therefore, the total number of trials was 480 (4 GIB layouts * 4 types of tasks * 30 trials). Each task had 120 trials, and we divided these trials into six blocks (20 trials per block). To avoid confusing the participants, only one kind of task was performed at a time for each of the different GIB layouts. In addition, because the participants could become accustomed to the data if we showed them the network diagrams which visualize the same data, we presented different data for each trial that was generated using the method explained in Section 3.3. We randomized the trial order for each task; therefore, the presented GIB layouts were randomly changed. The participants performed all four types of tasks in random order by considering the influence of familiarity and fatigue. Participants took a brief break of approximately 30 s after each block and a relatively long break after each task (6 blocks). The longest break was up to 5 min. Participants were instructed to accurately solve each problem. Time limits were not set for each problem, allowing the participants sufficient time to select the correct answers. If they focused on answering quickly, there was a possibility that we may encounter high error rates, which was not the intention of this study.

3.6. Experimental setup

The experiment was conducted in our laboratory, which was illuminated with artificial lighting. The task was displayed on a 24-inch monitor, with a resolution of 1920 × 1080 pixels.

3.7. Results

The results of the tasks are presented in Table 2. For task 1, the mean correct answer rate of all the layouts is 98.3%. It seems that participants had to count the number of groups in this task, and they answered correctly when they spent a considerable amount of time on this task. Therefore, the number of boxes is considered to be the only element affecting the correct answer rate, and this task can be considered to be a simple task. For task 2, the mean correct answer rate of all the layouts is 83.3%. The participants achieved a relatively high accuracy. In this task, the participants had to find the group with the largest number of nodes. It seems that there were no visualization factors that affected the correct answer rate except the area of the box, which was proportional to the number of nodes. Therefore, this task can also be considered to be a simple task. For task 3 and task 4, the mean correct answer rate of all layouts is 71.6% and 61.6%, respectively. The correct answer rate of task 4 was observed to be the lowest among all the tasks. We did not select task 1 and task 2 as the target of the experiment 2. Task 4 is a task related to the number of inter-edges and participants focus on inter-edges. Further, we did not select

Table 1
Parameters used for generating data.

m_{mean}	m_{stdev}	m_{min}	m_{max}	v_{mean}	v_{stdev}	v_{min}	p_{in}	p_{group}	p_{bridge}	p_{out}
11.4	5.4	6	17	21.0	14.12	4	0.0858	0.06	0.015	0.0006

Table 2
Results of user experiments with respect to mean accuracy (mean completion time).

	Task 1	Task 2	Task 3	Task 4
ST-GIB	98.1% (3.99 s)	89.6% (2.54 s)	67.4% (3.61 s)	62.8% (5.39 s)
CD-GIB	98.2% (4.53 s)	76.9% (2.76 s)	67.2% (3.84 s)	59.3% (5.59 s)
FD-GIB	98.7% (4.92 s)	82.9% (2.43 s)	78.8% (3.38 s)	59.3% (5.52 s)
TR-GIB	98.1% (4.49 s)	83.6% (2.71 s)	72.8% (3.85 s)	62.8% (5.13 s)
AVERAGE	98.3% (4.48 s)	83.3% (2.61 s)	71.6% (3.67 s)	61.6% (5.41 s)

task 4 as the target of the experiment 2 because it is difficult to control the difficulty level of the task and while analyzing the eye-tracking data, it is difficult to define AOIs for inter-edges. We selected task 3 as the target of the experiment 2 because, in this task, it is assumed that not only the number of intra-edges but also the area of the box and the density of intra-edges can affect the task performance. In this case, the density of intra-edges is the value obtained by dividing the number of intra-edges with the area of the circle that encloses all the nodes belonging to the group. In addition, the participants focused on the internal area of the box while performing this task. Thus, we can use the AOIs defined by the boxes. This task exhibits good compatibility with eye-tracking analyses. Further, we set two kinds of tasks for task 3, i.e., for selecting the group with the maximum and the minimum number of intra-edges. However, we selected the task of selecting the group with the maximum number of intra-edges to simplify the task. By investigating the result of task 3 in detail, we observed that FD-GIB, which looks considerably different from other layouts, produced the best result, followed by TR-GIB. According to the Wilcoxon signed-rank test, there was a significant difference in the task performance between the FD-GIB and TR-GIB (correct answer rate: $p = 0.017$; completion time: $p < 0.001$). Even in visualization involving same visualization elements, it is possible that the factors affecting the performance of the task can differ if the elements are arranged differently. This was the reason because of which these two layouts were selected. Further, in the experiment 2, we examined whether multiple exploration behaviors are present for each of these two layouts from the perspective of visual cognition.

4. Experiment 2

The objective of this experiment was to verify multiple exploration behaviors for the task that was selected in the experiment 1. We conducted a controlled laboratory experiment using the following two GIB layouts: FD-GIB and TR-GIB. Further, the participants were asked to perform a GIB evaluation task (i.e., which group has the maximum number of intra-edges). Subsequently, we investigated the exploration behavior with respect to visual cognition and correct answer rate. Herein, we discuss the experiment 2 in detail.

4.1. Data and layout generation

In this experiment, random data was generated using the same method as that used in the experiment 1. The setting is different from the one used in the experiment 1 (i.e., the number of groups). Further, we fixed the number of groups to 7 and 14.

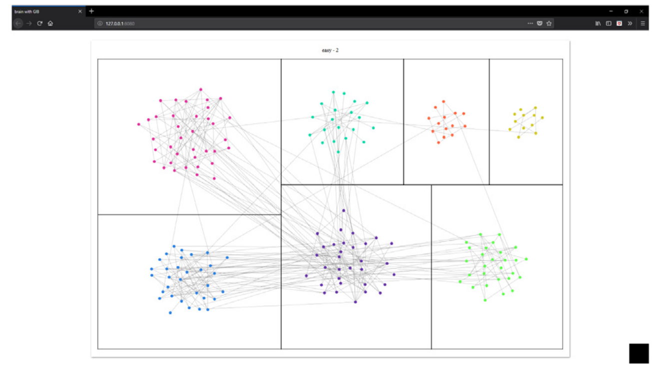


Fig. 2. A screenshot of the trial for TR-GIB in the easy level. Participants were asked to find the box with the largest number of intra-edges.

4.2. Participants

We used a within-subject study design with six participants. Five of the participants were male, and only one of the participants was female (age range: 21–32 years and mean: 25 years). Among the participants, one of the participants was familiar with GIB layouts, two participants were engaged in visualization study, and the rest were not familiar with GIB layouts but had adequate knowledge of the manner in which information can be extracted from diagrams and tables. All the participants had either normal or corrected-to-normal color vision. At the beginning of the experiment, the participants signed an informed consent form.

4.3. Study procedure

The experiments took 1.5–2 h, including the time for preparation, explanation and breaks. In the experiment, participants performed an intra-GNT task (i.e., which group has the maximum number of intra-edges?) for two types of GIB layouts, i.e., TR-GIB and FD-GIB. These tasks and target GIB layouts were selected from the results of the experiment 1. We recorded the eye movements of the participants to investigate the exploration behavior. Hence, at the beginning of the experiment, participants filled the informed consent form and the basic information questionnaire (name, age, and gender). After completing this step, we explained the eye-tracking system and each of the GIB layouts to the participants who were sitting at a position 65 cm from the screen. Next, we explained the task in detail, followed by a tutorial to ensure that the participants understood the network so that they can complete the tasks. Although the trial data in the tutorial was different from those in the actual experiments, the tutorial provided sufficient practical guide on the experimental procedures. Subsequently, the participants started the actual experiment. In the actual experiment, two difficulty levels were set for two types of GIB layouts, and 60 trials were prepared for each level. The difficulty level was set by the number of groups. In the easy trial, the number of groups was set to 7, while it was set to 14 in the difficult trial. Therefore, the total number of trials was 240 (2 GIB layouts * 2 difficulty levels * 60 trials). For each layout, the easy trials were divided into two blocks (30 trials per block) while the difficult trials were divided into three blocks (20 trials per block). Thus, the total number of blocks was 10. Because of the difference between the difficulty levels, the answer times

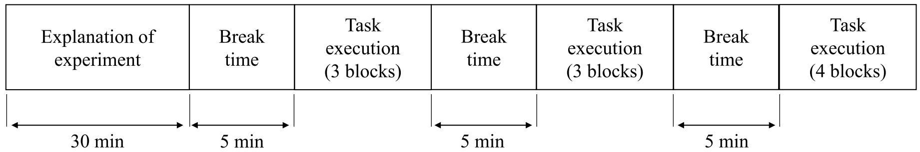


Fig. 3. Overview of the experiment.

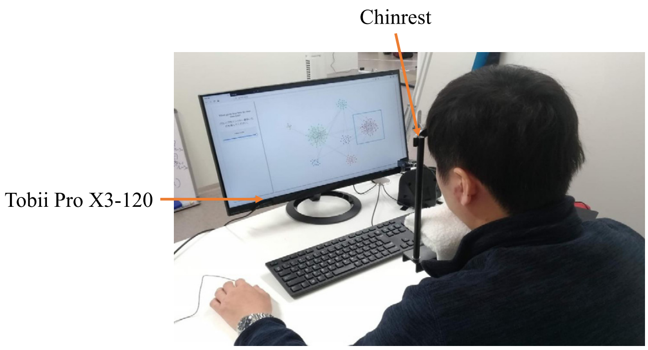


Fig. 4. Overview of the experimental setup. The task was displayed on a 24-inch monitor. The eye movements were tracked using a Tobii Pro X3-120 eye-tracking system fixed at the bottom of the display.

Fig. 5. A screenshot of the TR-GIB for the difficult level. The boxes with red edges indicate the AOIs. Each AOI is numbered in the descending order of the number of its intra-edges. In addition, we defined "in" and "out" AOI for AOI 1, AOI 2, and AOI 3.

per trial were different. Considering the influence of fatigue in the trials, the number of trials per block changed with respect to the difficulty level. Additionally, participants could become accustomed to the data if the network diagrams containing the same data are shown to them; therefore, we showed different data for each trial that was generated using the method explained in Section 4.1. Participants performed all the 10 blocks in a random order by considering the influence of both familiarity and fatigue. Participants took a brief break of approximately 30 s after each block and a relatively long break after three blocks. The longer break was up to 5 min, after which the eye-tracking system was recalibrated. Participants were instructed to solve each problem accurately. Time limits were not set for each trial, thereby allowing the participants sufficient time to choose the correct answers. If they focused on answering quickly, it would be possible to encounter high error rates and valueless eye-tracking data, which was not the intention of this study. An example of the trial is presented in Fig. 2, and an overview of the experiment is exhibited in Fig. 3.

4.4. Experimental setup

The experiment was conducted in our laboratory, which was illuminated with artificial lighting. The task was displayed on a 24-inch monitor, with a resolution of 1920 × 1080 pixels. The eye movements were recorded using the Tobii Pro X3-120 eye-tracking system. An overview of the experimental setup is presented in Fig. 4.

4.5. Data analysis

The eye-tracking data was preprocessed using Tobii Pro Studio (Tobii Pro Studio, 2018), an analysis software for eye-tracking data. We used an I-VT filter (Olsen, 2012) to determine whether a gaze was a fixation or a saccade.

After preprocessing the measured eye-tracking data, the following features were extracted. First, we defined the AOIs according to the boxes, and each AOI was numbered in descending order according to their number of intra-edges. We assumed that the participants consumed a lot of time to compare their answers, and we wanted to investigate the exploration behavior by which the participants were comparing the boxes; therefore, we divided the AOI 1, AOI 2, and AOI 3 into "in" and "out" using a circle that encloses all the nodes in the group, as depicted in Fig. 5. Therefore, in this analysis, AOI 1 is the correct answer, i.e., the box with the maximum number of intra-edges. Further, we calculated the fixation duration, fixation location based on the AOIs, and the fixation count for each AOI.

5. Result

The correct answer rate for the overall task performance was 78.1%, and the average completion time was 4.29 s. Although there were significant differences in task performance between the GIB layouts in the experiment 1, there were no significant differences observed in the experiment 2. It is considered that this is due to the difference in parameters regarding the number of boxes in the experiment 1 and this experiment.

5.1. Primary factor affecting the task performance

We assume that it is possible that a factor of visualization, which affects the task performance, can also be a factor related to exploration behavior. Hence, we investigated the visualization factors that affect the task performance. We believe that the following five factors may affect the task performance, especially the correct answer rate.

Factor 1 Number of boxes (7 or 14)

Factor 2 Whether the correct answer is the box with the largest area

Factor 3 The difference in the number of intra-edges between the correct answer and the box with the second most intra-edges

Factor 4 The number of inter-edges in the correct answer

(e) Transition matrix (FD-GIB, Correct Ans. = The largest box)

destination \ source	outside	AOI1	AOI2	AOI3	AOI4	AOI5	AOI6	AOI7
outside	-	23%	20%	17%	17%	30%	20%	28%
AOI1	33%	-	42%	43%	34%	32%	31%	24%
AOI2	19%	36%	-	18%	21%	14%	14%	13%
AOI3	14%	17%	14%	-	15%	11%	13%	8%
AOI4	10%	10%	10%	11%	-	8%	8%	13%
AOI5	9%	6%	6%	5%	7%	-	9%	7%
AOI6	7%	5%	4%	5%	4%	4%	-	6%
AOI7	8%	2%	4%	2%	3%	1%	4%	-

(g) Transition matrix (TR-GIB, Correct Ans. = The largest box)

destination \ source	outside	AOI1	AOI2	AOI3	AOI4	AOI5	AOI6	AOI7
outside	-	8%	8%	9%	12%	16%	9%	16%
AOI1	33%	-	50%	45%	32%	27%	31%	23%
AOI2	16%	40%	-	17%	13%	17%	20%	16%
AOI3	11%	22%	17%	-	20%	12%	10%	12%
AOI4	11%	13%	7%	13%	-	12%	10%	12%
AOI5	10%	7%	9%	7%	10%	-	11%	14%
AOI6	10%	6%	6%	5%	8%	10%	-	7%
AOI7	8%	3%	3%	4%	4%	6%	9%	-

(f) Transition matrix (FD-GIB, Correct Ans. ≠ The largest box)

destination \ source	outside	AOI1	AOI2	AOI3	AOI4	AOI5	AOI6	AOI7
outside	-	8%	32%	15%	13%	15%	20%	40%
AOI1	23%	-	21%	31%	23%	15%	20%	33%
AOI2	23%	41%	-	27%	31%	21%	13%	0%
AOI3	20%	18%	21%	-	26%	15%	23%	7%
AOI4	13%	15%	7%	12%	-	15%	0%	0%
AOI5	8%	12%	9%	9%	0%	-	20%	13%
AOI6	4%	5%	7%	7%	3%	18%	-	7%
AOI7	8%	1%	2%	0%	5%	0%	3%	-

(h) Transition matrix (TR-GIB, Correct Ans. ≠ The largest box)

destination \ source	outside	AOI1	AOI2	AOI3	AOI4	AOI5	AOI6	AOI7
outside	-	9%	5%	11%	6%	14%	26%	23%
AOI1	27%	-	44%	31%	26%	20%	18%	26%
AOI2	8%	42%	-	30%	15%	31%	19%	0%
AOI3	10%	27%	21%	-	31%	23%	11%	10%
AOI4	11%	8%	16%	15%	-	0%	18%	19%
AOI5	14%	7%	6%	6%	9%	-	5%	13%
AOI6	14%	5%	6%	4%	8%	11%	-	10%
AOI7	14%	1%	1%	1%	5%	0%	4%	-

Fig. 6. Result of comparing the eye-tracking data for conditions corresponding to whether the correct answer is the box with the largest area in FD-GIB and TR-GIB. Top row of ((a), (b), (c), and (d)) represent the example tasks while the second row of ((a), (b), (c), and (d)) represent the gaze plots for each participant. Each color corresponds to a different participant. The third row of ((a), (b), (c), and (d)) represent the heat maps of the fixation duration. ((e), (f), (g), and (h)) represent the probabilities of the gaze transitions between AOIs. Highlighted entries in the matrix in the same color belong to the same AOI pairs.

Factor 5 The difference in the density of intra-edges between the correct answer and the box with the second most number of intra-edges.

A multiple regression analysis was performed using the backward elimination method to determine whether these five factors considerably influenced the task performance. In this method, a regression equation was created using only independent variables. Subsequently, one independent variable with small influence was eliminated based on the t-value, indicating the influence of each independent variable in the regression equation. The criteria for eliminating an independent variable stated that the independent variable must have a minimum t-value of less than 2. Subsequently, we created a regression equation using one reduced independent variables and eliminated the independent variable again based on the t-value. This operation was performed until there was no independent variable to be eliminated. Finally, highly influential independent variables remained for the dependent variable. We used these five factors as independent variables, and the correct answer rate was used as the dependent variable. The results are presented in Tables 3 and 4 for FD-GIB and TR-GIB.

The results obtained denote that the correct answer rate in both layouts is considerably affected by factor 2, whether the correct answer is the box with the largest area (t-value = 8.41 for FD-GIB; t-value = 7.06 for TR-GIB). Further, based on the coefficients of each independent variable, we observed that the correct answer rate becomes higher when the number of boxes is small, the correct answer is the box with the largest area, and the difference in the number of intra-edges between the correct answer and the box with the second most intra-edges is large.

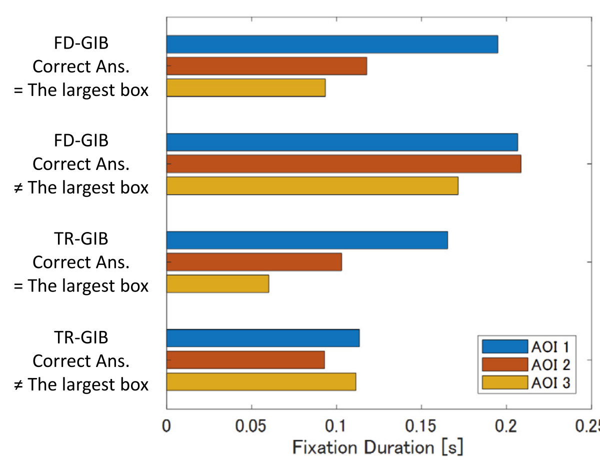


Fig. 7. Result of comparing the fixation duration in the AOI 1, AOI 2, and AOI 3 for the conditions corresponding to whether the correct answer is the box with the largest area in FD-GIB and TR-GIB.

Table 3
Result of multiple regression analysis for the correct answer rate in FD-GIB. ($R^2 = 0.558$, $R^2adj = 0.547$, $SEE = 0.181$).

Correct answer rate		Unstandardized coefficients		Standardized coefficients	t-value	p-value
		B	Std. error	Beta		
Variables	Constant	0.455	0.0759	–	5.99	0
	Factor1	−0.0225	0.00473	−0.079	−4.76	0
	Factor2	0.523	0.0622	0.523	8.41	0
	Factor3	0.00293	0.000613	0.077	4.78	0

Table 4
Result of multiple regression analysis for the correct answer rate in TR-GIB. ($R^2 = 0.543$, $R^2adj = 0.531$, $SEE = 0.179$).

Correct answer rate		Unstandardized coefficients		Standardized coefficients	t-value	p-value
		B	Std. error	Beta		
Variables	Constant	0.512	0.066	–	7.75	0
	Factor1	−0.0131	0.00466	−0.046	−2.8	0.005
	Factor2	0.347	0.0492	0.347	7.06	0
	Factor3	0.00407	0.000732	0.107	5.57	0

5.2. Analysis of the exploration behavior

As discussed in Section 4.5, our eye-tracking data analysis was based on the AOIs that were defined according to the property of groups, i.e., the number of intra-edges. Fig. 6 presents the result of comparing the eye-tracking data for the conditions in which the correct answer was or was not the box with the largest area. The top row in Fig. 6(a), (b), (c), and (d) depicts an example of FD-GIB and TR-GIB with seven boxes. The box with the most intra-edges is the box with red edges, followed by the box with blue edges and the box with green edges. Further, the box with the largest area is the box with a red face, followed by the box with a blue face and the box with a green face. Therefore, the box with red edges and red face is the box with both the maximum number of intra-edges and the largest area. It is worthy of note that these colors are purely explanatory; there were no color edges or faces in the experiment. The second row in Fig. 6(a), (b), (c), and (d) represents the gaze plots for each of the participants. The third row in Fig. 6(a), (b), (c), and (d) presents the heat maps of the fixation duration.

To investigate the exploration behavior based on the fixation location, we calculated the transition matrix of the AOIs (Fig. 6(e), (f), (g), and (h)). Each column indicates the relative amount of transition between a given AOI to any other AOI. In this analysis, we consider the boxes labeled AOI 8 to AOI 14 as outside because, in this analysis, we ignored the number of boxes that did not have considerable influence on the correct answer rate. Accordingly, this analysis ignored many transitions that began from and ended outside. However, the primary objective of this analysis is to identify the main differences in the exploration behavior between the two conditions, i.e., whether the correct answer is the box with the largest area. Therefore, our approach can be considered to be appropriate. In the future, we plan to analyze the eye-tracking data in greater detail.

By comparing the transition matrices, we obtained the following insights. In both the layouts, the transition probabilities of AOI 2 to AOI 3 (and vice versa) are higher when the correct answer is not the box with the largest area. Further, in both the layouts, the transition probabilities of AOI 1 to AOI 2 and AOI 3 are higher, and the probabilities of AOI 2 and AOI 3 to AOI 1 are lower than when the correct answer is the box with the largest area.

Based on this result, the difference and the average fixation duration for AOI 1 to AOI 3 were calculated. Fig. 7 depicts the result of the comparison of the fixation duration when the correct answer either is or is not the box with the largest area in both the layouts. The fixation duration for each AOI revealed that in both the layouts, the time spent on AOI 1 was the longest when the correct answer is the box with the largest area. However, when the correct answer is not the box with the largest area, the time spent on AOI 2 was longer than that the time spent on AOI 1 for FD-GIB, whereas the time spent on AOI 3 was as long as that on AOI 1 for TR-GIB.

In summary, participants tend to frequently compare and focus on AOIs other than AOI 1 when the correct answer is not the box with the largest area. This indicates that it is possible that the participants may tend to focus on the size of boxes rather than the number of intra-edges.

We analyzed the location at which the participants focused on when they were comparing the AOIs. The difference between the exploration behaviors because of the difference in conditions appeared mainly in the eye trajectory data for AOI 1, AOI 2, and AOI 3; thus, we further analyzed the eye-tracking data for those AOIs by dividing these AOIs into "in" and "out". If the participants focus on the number of intra-edges, the fixation duration for "in" will increase. Further, if the participants focus on the size of the boxes, the fixation duration for "out" will also increase. We calculated the ratio of fixation duration for "in" and "out" because the time taken for one trial depended on the participants. We classified the eye-tracking data relating to AOI 1, AOI 2, and AOI 3 based on "in" and "out". Subsequently, we calculated the ratio of fixation duration for "in" to the sum of fixation duration for "in" and "out". For both the layouts, we compared the ratio for conditions in which the correct answer was or was not the box with the largest area. The result is presented in Fig. 8. Wilcoxon signed-rank test indicated the absence of any significant difference between the two conditions. Based on this result, it was revealed that the participants tended to fix their eyes on the "in" area regardless of the condition of the size of the boxes.

In summary, we discovered that whether the correct answer is the box with the largest area, which is a visualization factor that considerably affected the correct answer rate, changed the exploration behavior; however, the fixation point for the boxes was not changed owing to the change in the size of the boxes. It seems that participants focused on the number of edges, even though they can obtain information related to the boxes using peripheral vision and this information changed their exploration behavior.

6. Discussion

A visualization method is based on human visual perception. Therefore, we need to know the human behavior when they use the visualization system to advance our understanding of visualization and the design of interactions and visual analytics system considering the human factors. In this study, we investigated

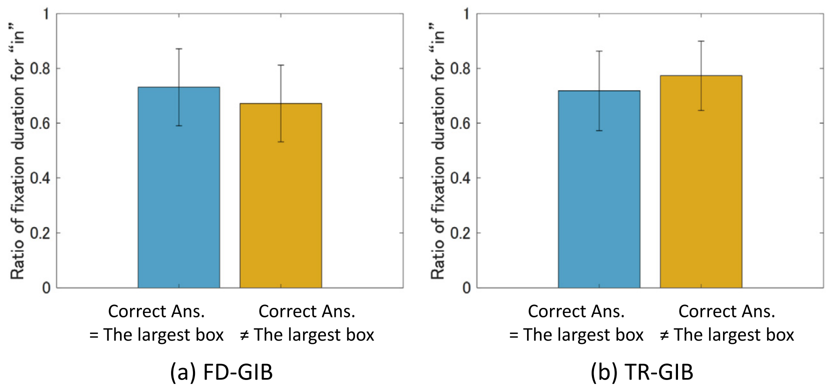


Fig. 8. Result of comparing the ratio of fixation duration and fixation count for the condition corresponding to whether the correct answer is the box with the largest area.

the exploration behavior for the complicated task of GIB layouts. There are few studies investigating the exploration behavior for complicated visualization evaluation task.

In this study, six participants performed the exploration task using GIB layouts. The task (i.e., which group has the maximum number of intra-edges?) was coupled with the FD-GIB and TR-GIB layouts. The experimental results revealed that the correspondence of the box with the maximum number of intra-edges to the box with the largest area considerably affects the correct answer rate. The eye-tracking study confirmed that in both FD-GIB and TR-GIB, the exploration behavior changed according to the visualization factor affecting the correct answer rate. However, the fixation point of the boxes did not change depending on the visualization factors. From these results, we observed that the participants focused on the appropriate visualization factor to complete the task, i.e., the number of intra-edges; however the exploration behavior was affected by the conditions corresponding to whether the correct answer is the box with the largest area which affected the correct answer rate.

In this experiment, we set the feature of the visualization elements based on random parameters except for the number of boxes. It was revealed that the condition of the size of the boxes affected the task performance, even though we could not quantitatively analyze the influence of differences in the area of the box with the correct answer and the box with the second most intra-edge on the correct answer rate because of random parameters. Therefore, our future work would focus on analyzing in more detail the effects of the size of box on the correct answer rate.

7. Conclusion

This study made the following two main contributions: (1) an evaluation of the visualization factors affecting the task performance and (2) an examination of the exploration behavior with respect to the visualization element.

The obtained result suggest that the visualization elements that are not intended for the visualization designer can influence the task of extracting information from the data. Therefore, the designers need to configure the visualization by considering the human behavior and visual cognition conditions. To avoid the influence from unintended visualization factors and subsequently improve the design of GIB layout, we propose to add user interaction, in which user can select appropriate visualization method according to the information the user wants to know. As an example, when a user wants to know about the number of intra-edges, it is suggested that to color the nodes of each group according to the number of edges is useful for a user.

Acknowledgments

This work was supported by JST CREST, Japan Grant Number JPMJCR1511, Japan, and The Keihanshin Consortium for Fostering the Next Generation of Global Leaders in Research (K-CONNEX), established by Human Resource Development Program for Science and Technology, MEXT.

References

Andrienko, G., Andrienko, N., Burch, M., Weiskopf, D., 2012. Visual analytics methodology for eye movement studies. IEEE Trans. Vis. Comput. Graphics 18 (12), 2889–2898. http://dx.doi.org/10.1109/TVCG.2012.276.

Becker, R.A., Eick, S.G., Wilks, A.R., 1995. Visualizing network data. IEEE Trans. Vis. Comput. Graphics 1 (1), 16–28. http://dx.doi.org/10.1109/2945.468391.

Blascheck, T., John, M., Kurzhals, K., Koch, S., Ertl, T., 2016. VA 2: A visual analytics approach for evaluating visual analytics applications. IEEE Trans. Vis. Comput. Graphics 22 (1), 61–70. http://dx.doi.org/10.1109/TVCG.2015.2467871.

Bruls, M., Huizing, K., Van Wijk, J.J., 2000. Squarified treemaps. In: Data Visualization 2000. Springer, pp. 33–42. http://dx.doi.org/10.1007/978-3-7091-6783-0_4.

Burch, M., Konevtsova, N., Heinrich, J., Hoeferlin, M., Weiskopf, D., 2011. Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. IEEE Trans. Vis. Comput. Graphics 17 (12), 2440–2448. http://dx.doi.org/10.1109/TVCG.2011.193.

Chaturvedi, S., Dunne, C., Ashktorab, Z., Zachariah, R., Shneiderman, B., 2014. Group-in-a-box meta-layouts for topological clusters and attribute-based groups: Space-efficient visualizations of network communities and their ties. Comput. Graph. Forum 33, 52–68. http://dx.doi.org/10.1111/cgf.12400.

Gansner, E.R., Hu, Y., 2008. Efficient node overlap removal using a proximity stress model. In: International Symposium on Graph Drawing. Springer, pp. 206–217. http://dx.doi.org/10.1007/978-3-642-00219-9_20.

Huang, W., Eades, P., Hong, S.-H., 2008. Beyond time and error: a cognitive approach to the evaluation of graph drawings. In: Proceedings of the 2008 Workshop on beyond Time and Errors: Novel EvaLuation Methods for Information Visualization. ACM, p. 3.

Kim, S., Dong, Z., Xian, H., Upatising, B., Yi, J.S., 2012. Does an eye tracker tell the truth about visualizations?: Findings while investigating visualizations for decision making. IEEE Trans. Vis. Comput. Graphics 18 (12), 2421–2430. http://dx.doi.org/10.1109/TVCG.2012.215.

Kobourov, S.G., 2004. 12 force-directed drawing algorithms.

Netzel, R., Burch, M., Weiskopf, D., 2014. Comparative eye tracking study on node-link visualizations of trajectories. IEEE Trans. Vis. Comput. Graphics 20 (12), 2221–2230. http://dx.doi.org/10.1109/TVCG.2014.2346420.

Netzel, R., Hlawatsch, M., Burch, M., Balakrishnan, S., Schmauder, H., Weiskopf, D., 2017. An evaluation of visual search support in maps. IEEE Trans. Vis. Comput. Graphics 23 (1), 421–430. http://dx.doi.org/10.1109/TVCG.2016.2598898.

Olsen, A., 2012. The tobii i-vt fixation filter, Tobii Technology.

Onoue, Y., Koyamada, K., 2017. Optimal tree reordering for group-in-a-box graph layouts. In: SIGGRAPH Asia 2017 Symposium on Visualization. In: SA '17, ACM, New York, NY, USA, pp. 13:1–13:9. http://dx.doi.org/10.1145/3139295.3139308.

Pohl, M., Schmitt, M., Diehl, S., 2009. Comparing the readability of graph layouts using eyetracking and task-oriented analysis. In: Proceedings of the Fifth Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging. In: Computational Aesthetics'09, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, pp. 49–56. http://dx.doi.org/10.2312/COMPAESTH/COMPAESTH09/049-056.

Purchase, H., 1997. Which aesthetic has the greatest effect on human understanding?. In: International Symposium on Graph Drawing. Springer, pp. 248–261. http://dx.doi.org/10.1007/3-540-63938-1_67.

Purchase, H.C., 1998. Performance of layout algorithms: Comprehension, not computation. J. Visual Lang. Comput. 9 (6), 647–657. http://dx.doi.org/10.1006/jvlc.1998.0093.

Purchase, H.C., Carrington, D., Allder, J.-A., 2002. Empirical evaluation of aesthetics-based graph layout. Empir. Softw. Eng. 7 (3), 233–255. http://dx.doi.org/10.1023/A:1016344215610.

Rodrigues, E.M., Milic-Frayling, N., Smith, M., Shneiderman, B., Hansen, D., 2011. Group-in-a-box layout for multi-faceted analysis of communities. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, pp. 354–361, http://dx.doi.org/10.1109/PASSAT/SocialCom.2011.139.

Saket, B., Simonetto, P., Kobourov, S., 2014. Group-level graph visualization taxonomy. arXiv preprint arXiv:1403.7421.

Tobii Pro Studio, 2018. https://www.tobiipro.com/product-listing/tobii-pro-studio/, (accessed on Decenber 11, 2018).

Vehlow, C., Beck, F., Weiskopf, D., 2017. Visualizing group structures in graphs: A survey. In: Computer Graphics Forum, vol. 36(6). Wiley Online Library, pp. 201–225, http://dx.doi.org/10.1111/cgf.12872.